Freehand Gestural Text Entry for Interactive TV

Gang Ren
Department of Computer Science
University of Bath
Bath, UK, BA2 7AY
garry.ren@gmail.com

Eamonn O'Neill
Department of Computer Science
University of Bath
Bath, UK, BA2 7AY
eamonn@cs.bath.ac.uk

ABSTRACT
Users increasingly expect more interactive experiences with TV. Combined with the recent development of freehand gestural interaction enabled by inexpensive sensors, interactive television has the potential to offer a highly usable and engaging experience. However, common interaction tasks such as text input are still challenging with such systems. In this paper, we investigate text entry using freehand gestures captured with a low-cost sensor system. Two virtual keyboard layouts and three selection techniques were designed and evaluated. Results show that a text entry method with dual circle layout and an expanding target selection technique offers ease of use and error tolerance, key features if we are to increase the use and enhance the experience of interactive TV in the living room.

Categories and Subject Descriptors
H.5.2 [**Information interfaces and presentation**]: User Interfaces – *Input devices and strategies, Interaction styles.*

General Terms
Human Factors; Design; Measurement.

Keywords
Text entry; Freehand gesture; Expanding Target.

1. INTRODUCTION
The increasing use of interactive TV in the living room brings novel opportunities and requirements for rich and engaging interactive experiences. There are many different ways to interact with an interactive TV. The most common input method is using a traditional remote, however, the remote normally offers only a limited set of buttons and does not lend itself to offering richer means of interaction. Other input methods used for computers or mobile devices can also be used with interactive TVs, such as keyboard, mouse, and touch-sensitive displays. However, these input devices are usually installed close to or even contiguous with the screen so they are not suitable for typical use scenarios with an interactive TV where the user is often at a distance from the screen. Mobile devices such as phones or tablets can also be used to interact with remote displays, but configuration is often needed to connect the personal devices so this may not be convenient in some scenarios.

Gestural input is increasingly popular, using hands-on input devices (e.g. Wii Remote) or freehand motion tracking by a camera (e.g. Microsoft Kinect). As gestural input moves beyond home gaming settings, freehand gestural interaction, which has no need for hands-on input devices and so enables easier and more convenient "walk up and use" [3], is likely to become more important in interactions with TV in everyday settings.

Currently, however, it is still difficult to perform some common tasks such as text entry with freehand gestural interaction. For example, when a person is trying to search for a program or a video clip on an interactive TV, her text entry task may be challenging due to several factors including, for example, the relatively low resolution of many remote gesture sensors and the distance to the TV screen.

Although research has been conducted on text input with various input devices and techniques, most techniques use handheld input devices and so cannot be used directly in freehand interaction. Therefore, we are motivated to investigate freehand gestural text input methods. Here, we report findings from the design and evaluation of some candidate virtual keyboard layouts and input techniques.

2. RELATED WORK
2.1 Gestural Interaction and Interactive TV
Gestural interaction has been investigated for a long time and many different gesture types have been designed and evaluated, and efforts made to summarize and classify different types of gesture [18, 37, 39]. Karam and Schraefel [18] classified gesture styles as deictic, manipulation, gesticulation, semaphores and sign language. Deictic and manipulative gestures are similar to pointing and manipulating in real life interaction, and they could be used without special learning or training. Gesticulation is the gesturing that accompanies everyday speech so it too requires no special training. In contrast, semaphoric gesture and sign language require a dictionary and even grammatical structures, thus training is necessary before using these gesture types with an interactive system.

Various input devices and gestures have been investigated with interactive TV. Bobeth et al. [5] tested freehand menu selection for interactive TV with 4 different designs, and found that freehand gestures could be an appropriate way for older adults to control a TV. A selection task was also investigated in [29], and participants preferred freehand gestural pointing to using a hand-held pointing device. Drawing different shapes in the air can also be used to select objects or menu items [2] with interactive TV, however, certain shapes are not easy to perform and remember, and have low recognition rates. User defined gestures for TV were also evaluated in [38]. The results showed that a pointing action was frequently used and a desktop interaction style, such as a push in mid-air to simulate clicking, was observed in many cases.

EuroITV'13, June 24-26, 2013, Como, Italy.

However, there has been some criticism of semaphoric and sign language style gestural interaction, including arguments that gesture is a step backwards to command-line interfaces [27], natural user interfaces are not natural [26], and that gestural interaction should be based on well designed metaphors rather than gesture design [16].

2.2 Text Entry and Interactive TV

The handheld remote control is by far the main input device for TVs, and many text input methods designed for TV remotes have been proposed and investigated [7, 14, 35]. Geleijnse et al. [9] also compared the physical Qwerty keyboard and remote control for text entry and suggested that the Qwerty keyboard is better.

Gestural text input methods have gained more research interest recently and many gestural text entry methods have been proposed and investigated. Most of these methods could also be used with interactive TV. For example, Jones et al [17] used accelerometer-based gesture enabled by a Wii remote and virtual keyboard for text entry. Users achieved 3.7 words per minute (wpm) in first time use and 5.4 wpm after 4 days' practice. A stroke-based text entry method was also designed with data gloves and fiducial markers [25]. Users reached 6.5 wpm without word completion after 2 weeks' practice. A Wii remote was used for text input with large displays in [34]. Three different layouts – circle, Qwerty and 3D cube – were evaluated and the Qwerty layout had the best performance (18.9 wpm), but decreased significantly with more errors as the user moved away from the display. Kristensson and Zhai [20] investigated shape writing recognition to perform word-based text input with a stylus keyboard, and saw high performance in informal trials. However, a test of text entry methods on mobile touch screens showed Qwerty was faster than handwriting and shape writing text entry, and handwriting was the slowest and least accurate text entry technique [6]. Other text entry methods originally designed for stylus and touch screen, such as FlowMenu [12] or Quikwriting [15, 28] could also be used with freehand gestural text input.

Besides holding a tracked device in the hand (e.g. Wii remote) or wearing a data glove or fidual markers, it is also possible to track freehand gesture with low-cost remote cameras, such as Microsoft Kinect[1] or ASUS Xtion[2]. This type of tracking device has the advantage of enabling freehand tracking in 3D space without requiring the user to hold any device in the hand or use fiducial markers. However, such tracking techniques with a single remote camera normally have low resolution and tracking accuracy. For example, the accuracy of the Microsoft Kinect depth sensor is about 3mm in the image plane and about 1cm in depth at a distance of 2 meters [30, 40]. In practice, the skeleton tracking based on the raw depth data can be even noisier.

Kristensson et al. [19] investigated freehand text entry using freeform alphabetic character recognition, and the evaluation shows a recognition accuracy of 92.7%–96.2%, however, no evaluation on text entry performance is available from their study. Freehand gesture was also used with speech recognition for text entry [13], and 5.18 wpm text input speed was achieved. Although freehand gestural text input with a virtual keyboard is widely used in commercial products (such as games designed for Microsoft Kinect), there is very little research available on this topic. A previous study showed that with the default Xbox 360 gesture based text input interface, the input speed was only 1.83 wpm [13].

Text input with a virtual keyboard is basically a sequence selection of small targets packed together on the keyboard. This is a challenging task for noisy motion tracking with a low-cost single camera system. For example, one typical issue for freehand gesture text input is the difficulty of gesture delimiter design [3]. Accot and Zhai proposed a cross technique which can be used for selection without clicking [1]. A similar method was used for freehand selection in [31], in which users reach towards the target to select without the need to stay inside the target. Furthermore, although an "expanding target" can be used to support selection for small targets, it can only magnify in visual space but not in motor space when targets are closely packed [24]. Although a predictor could be used to increase the motor space before the cursor enters the target area, the benefit is very limited [24].

3. DESIGN OF FREEHAND TEXT ENTRY

3.1 Design Considerations

Our aim is to implement text entry methods that facilitate "walk-up-and-use" – or in the case of interactive TV, more likely "sit-down-and-use" – interaction experiences [3], while retaining the simplicity and directness of freehand interaction. Most previous gestural text entry methods use handheld devices or fiducial markers for tracking motion, which can offer accurate tracking of hand, wrist and fingers. Freehand motion tracking enabled by an inexpensive remote camera, on the other hand, can track hand motion robustly but not the small motions of wrists or fingers, especially when users are at a distance from the display/sensor. Besides the lack of fine movement tracking, there are also some other challenges for freehand text input, such as noisy motion tracking, no physical button or surface to click or touch, and no physical support or tactile feedback for the hand.

Text entry methods based on freeform alphabetic character recognition have been investigated in considerable previous work [e.g. [19, 25, 41]. With such methods, however, users need to learn and remember a set of gestures. Such learning demands may not be suitable for scenarios with interactive TV where quick and easy interaction is important. Text entry methods based on word level prediction, such as shorthand writing [20, 43], Swype[3] or text input based on speech recognition [13] are also possible for freehand text entry. But with interactive TV applications, the requirements of non-dictionary word entry could be high (e.g. entering user name, password, email address, or url), thus character based text input may be better suited to interactive TV.

A virtual keyboard can provide easy recognition and learning [17, 34] and, therefore, may be more suitable for interactive TV text entry. Although character arrangement on a virtual keyboard can be optimized according to the context of use and alternative arrangements may improve the performance of expert users [4, 12, 15, 17, 33], the Qwerty layout still has some benefits [34, 42], is the basis of many improved text entry methods [8, 21, 42] and has the advantage of familiarity to many users. For "walk up and use" and entertainment scenarios with interactive TV, the familiarity of the Qwerty layout is very important to users, with less demand for extra learning and less visual scan time. Thus we designed our gestural text entry based mainly on the Qwerty keyboard layout and a character based text entry method.

[1] http://www.microsoft.com/en-us/kinectforwindows/

[2] http://www.asus.com/Multimedia/Xtion_PRO/

[3] http://www.swype.com/

One benefit of freehand gestural interaction is that the hand can move in 3D space, which means that the virtual keyboard can be in 3D. However, results from previous research [34] indicate that 3D layout text entry has low performance. And previous work on 2D and 3D option selection with freehand gesture [32] also suggests that freehand selection with a 3D layout is less accurate than with a 2D layout. Thus, we designed and evaluated a 2D keyboard layout in this study.

3.2 Keyboard Layout

3.2.1 Qwerty

As noted above, Qwerty is a very familiar keyboard layout and has been shown to perform well as a virtual keyboard with a mid-air handheld device [34]. It is therefore a reasonable candidate for freehand text entry. For our prototype design and evaluation, we used a Qwerty layout of 28 characters (26 English letters, space and backspace), similar to the keyboard used for touch-screens such as Windows Phone (Figure 1). More keys could also be added in different applications.

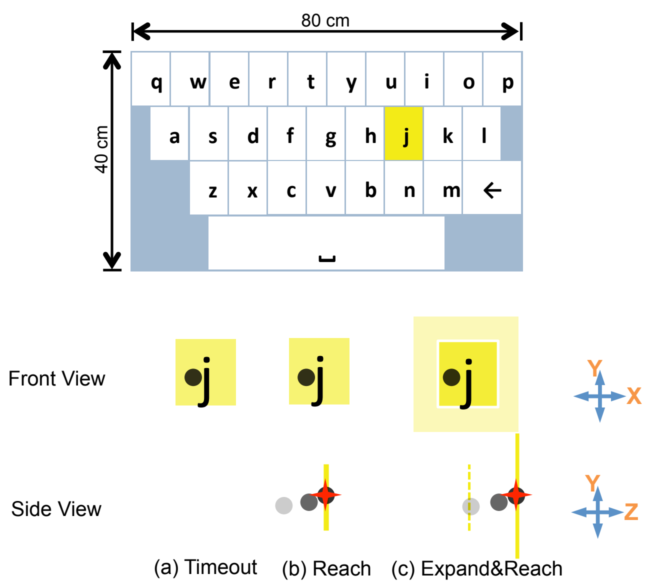


Figure 1. Up: Qwerty layout. Down: Selection techniques; the spherical grey cursor is controlled by the user's hand position.

3.2.2 Dual-circle

Besides the Qwerty layout, another virtual keyboard layout that has been investigated is circle. Although the circle layout is not as effective as Qwerty with a mid-air handheld device [34], it may bring benefits for freehand interaction. For example, rather than being tiled, characters are arranged to offer easy access to each character from the center of the circle. With accurate handheld devices such as in [34], all characters could be distributed in a single circle. However, for freehand motion tracked by low cost camera, the character size could be too small for reliable use with noisy tracking input in an interactive TV scenario.

To address this issue and to leverage the two-handed operation that freehand gestural interaction allows, we proposed a Dual-circle layout for text entry (Figures 2 and 3). The characters are evenly distributed in 2 circles next to each other. Thus each character can be bigger than if they were distributed in one similarly sized circle. To leverage users' familiarity with the Qwerty layout, we based the character distribution on Qwerty: the top and bottom of each circle is used for characters located in the top row and bottom row in Qwerty, and the middle row of the Qwerty layout is turned vertically and put sideways in each circle based on the corresponding hand and fingers when using a Qwerty keyboard. Gaps are used to separate the different character groups for a clearer mapping to the familiar Qwerty layout.

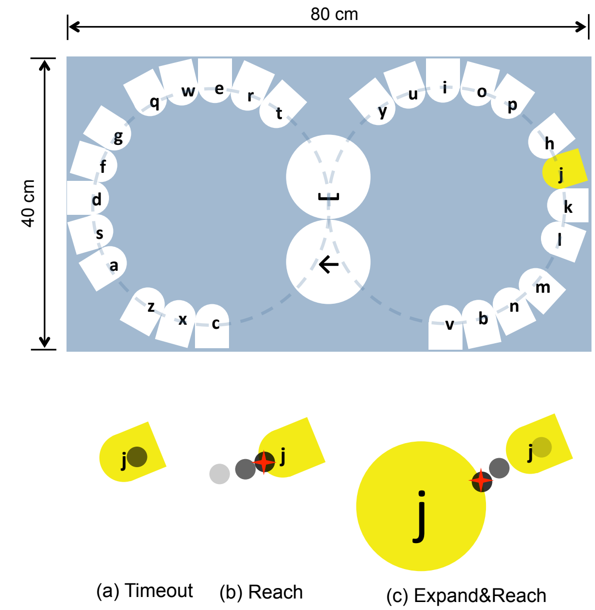


Figure 2. Up: Dual-circle layout. Down: Selection techniques; spherical grey cursor controlled by the user's hand position.

Space and backspace are represented by circles located in the middle of the keyboard for easy access with both hands. As users dislike selecting in the left-down direction with the right hand [31], the left-down portion of the right circle and the mirrored portion of the left circle are left blank.

3.3 Character Selection

As noted above, without a physical device in the hand and buttons to click, freehand gestural selection can be challenging. And although finger and wrist movement are used in some previous work, they are less suited to freehand interaction tracked by a remote inexpensive camera, so other techniques are required.

3.3.1 Timeout

One common method for freehand selection is pointing to the target and waiting for a timeout threshold. It is easy for novices to understand and perform and is accurate for large targets. The primary disadvantage is that the dwell time can slow overall selection time. For both our layouts, the timeout selection method can be used: the user points to a character by the X-Y position of her hand and waits for the timeout threshold, e.g. 1.2 s, to select the character (Figure 1.a, Figure 2.a).

3.3.2 Reach

As an alternative to timeout, hand motion can also be used for selection confirmation. The Reach technique has been used for target selection with freehand gestures [31], in which users select by moving their hands to reach into the target in 3D. For a virtual Qwerty keyboard placed vertically in front of the user's body position (in this case at 40 cm), the user can point to the desired character by X-Y movement of the hand and then reach forward in the Z-dimension to select the character (Figure 1.b). For the Dual-circle layout, the character can be selected by moving the hand's X-Y position to reach across the border of the desired character tab (Figure 2.b).

With the Reach technique, although 3D hand position is required in the Qwerty layout while only X-Y position is required in the Dual-circle layout, with both layouts the user's hands move freely in 3D space to reach the characters. In practice, with both Qwerty and Dual-circle layouts, the user tends to move the selecting hand forward and towards the target character simultaneously in one fluid movement.

3.3.3 Expand&Reach

In both layouts, the character size is relatively small so could be difficult and error-prone for freehand selection. It is also difficult to expand the target in motor space for tiled targets in 2D interfaces [24]. However, since the hand can move in 3D space, the combination of the extra dimension and the Reach selection technique brings new interaction opportunities. We designed additional Expand&Reach techniques for both keyboard layouts. With the Qwerty layout, when the user points to a character, an expanded target appears along the Z-dimension (e.g. 5 cm further away than the current hand position and 2 times bigger than original size). The user moves her hand forward to reach the expanded target in order to select (Figure 1.c). With the Dual-circle layout, when the user points to a character, an expanded character tab containing the target character appears in the center of the corresponding circle (Figure 3). The user moves her hand to reach the expanded target to select it (Figure 2.c).

There are several potential advantages of the Expand&Reach technique: (i) easier selection – the target expands in both visual and motor space; (ii) error tolerance – users need to move their hand to reach the expanded target to confirm selection, so if they notice a selection error before reaching the expanded target, they have a chance to (re)select the right character; (iii) requires only hand position tracking – no fine finger movement or posture tracking is needed.

4. EXPERIMENTAL EVALUATION

A controlled experimental evaluation was conducted to investigate the effects of the different keyboard layouts and selection techniques.

4.1 Independent variables

The independent variables were Layout (Qwerty, Dual-circle), Selection Techniques (Timeout, Reach, Expand&Reach), and Day (1 to 5).

4.2 Participants

6 participants (4 males, 2 females) were recruited from the local campus, mean age 26 (sd = 1.7), all right handed and with some experience of gestural interaction for gaming.

4.3 Procedure

The evaluation lasted for 5 days with 6 sessions every day. Each session tested a combination of Layouts and Selection techniques. In each session, 4 sentences were presented for the participant to reproduce, the first as practice followed by 3 test sentences. Six sets of sentences were randomly selected from MacKenzie and Soukoreff's phrase sets [22] and were assigned randomly to different sessions. User preferences and NASA task load index (TLX) data were collected on the first and last days.

Each character is white and highlights yellow when pointed at. With Timeout selection, the color gradually changes from yellow to green until timeout. Using Reach and Expand&Reach with Qwerty, the character gradually changes from yellow to green as the hand moves forward to reach it. The selected character appears immediately below the target sentence. A "typing" sound is played when a character is entered correctly. If the entry is incorrect, an error sound is played instead and the input is shown in red. All mistakes must be corrected for each sentence.

Both keyboards were 80 cm x 40 cm, with the top edge at the same height as the user's shoulders in motor space. Two spheres sized 2 cm were controlled with the hands. With the Qwerty layout, the hand in front of the other is enabled and rendered in black, the other is rendered grey and disabled to avoid accidental selection. With the Dual-circle layout, the blank center area is large enough to accommodate an idle hand without accidental selection, so no disable mechanism was used. For timeout selection technique with both layouts, the dwell time was 1.2 s. The 1.2 s dwell time was based on a pilot study which showed that less than 1.2 s produced more errors, and the observation that almost all commercial Kinect interfaces with timeout selection use more than 1.5 s.

4.4 Experimental Setting

A Sanyo PDG-DWL2500 3D projector was used at 1280 x 720 resolution with a 203 x 115 cm screen centered at 130 cm height to simulate a large interactive TV display. A Microsoft Kinect camera was used with a refresh rate of 30 fps and the Kinect for Windows SDK V1.5 on Windows 7. The Kinect camera was placed 50 cm in front of the screen at a height of 70 cm. The user stood 250 cm from the screen (Figure 3).

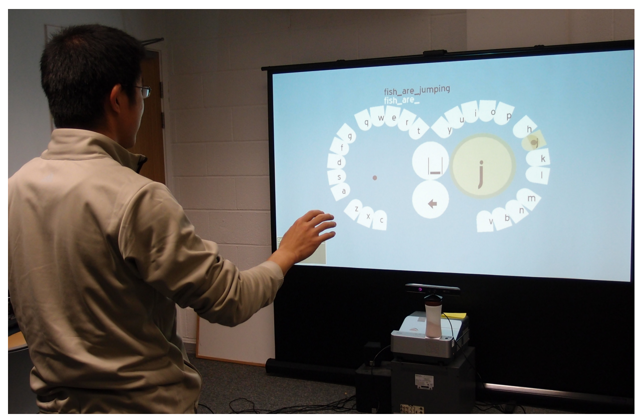


Figure 3. Experimental setting.

5. RESULTS

5.1 Typing Speed

A repeated-measures ANOVA for Layout x Selection Technique x Day was used to analyze the text input speed. Main effects were found for Selection Technique ($F_{2,10}$=144.27, p<.001) and Day ($F_{4,20}$=21.49, p<.001). Layout had no significant effect ($F_{1,5}$=1.38, p=.29). Interaction effects were found for Layout x Selection Technique ($F_{2,10}$=11.69, p<.01) and Day x Selection Technique ($F_{8,40}$=6.05, p<.001).

Post hoc Bonferroni pairwise comparisons showed that Reach and Expand&Reach were both significantly faster than Timeout (p<.001), with no significant difference between them. Text input speed in the last 3 days was significantly faster than on the first day (p<.05), with no significant difference between the last 3 days. Mean text input speeds across all conditions are shown in Figure 4 and Table 1.

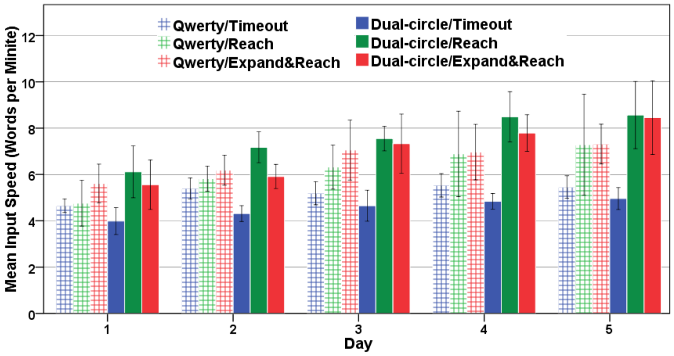


Figure 4. Mean text input speed. In this and later charts, error bars represent 95% confidence intervals.

Table 1. Mean speed (wpm) in day 1, day 5 and over 5 days.

	Qwerty			Dual-circle		
	Time-out	Reach	Expand &Reach	Time-out	Reach	Expand &Reach
Day 1	4.65	4.76	5.61	3.99	6.11	5.56
Day 5	5.46	7.29	7.31	4.96	8.57	8.46
5 days overall	5.25	6.22	6.63	4.55	7.58	7.01

5.2 Error Rate

A repeated-measures ANOVA for Layout x Selection Technique x Day was used to analyze error rate. Main effects were found for Layout ($F_{1,5}$=10.86, p<.05) and Selection Technique ($F_{2,10}$=11.09, p<.01). Day had no significant effect ($F_{4,20}$=1.94, p=.14). No interaction effect was found. Mean error rates are shown in Figure 5 and Table 2.

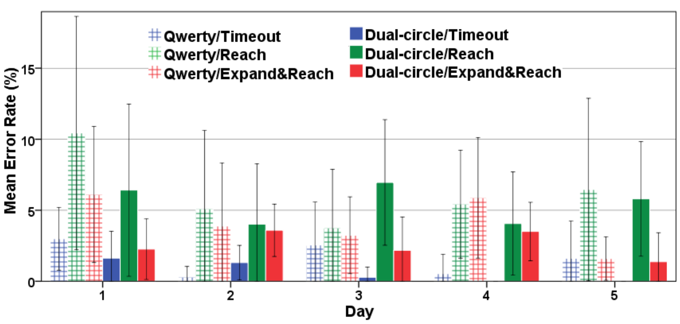


Figure 5. Mean error rate.

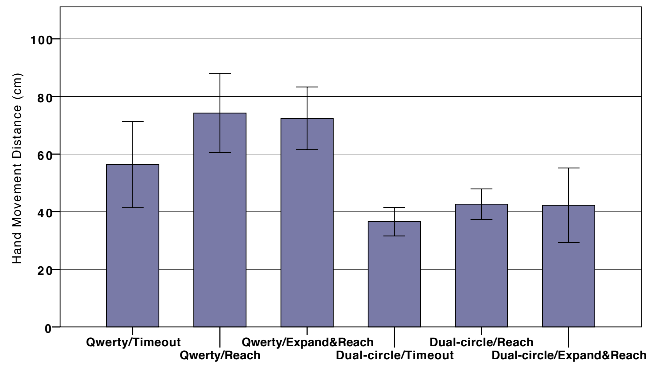
Post hoc Bonferroni pairwise comparisons showed Qwerty to be significantly more error-prone than Dual-Circle (p<.05). Timeout had significantly fewer errors than Reach (p<.05). There were no other significant differences between selection techniques.

Table 2. Mean error rate (%) in day 1, day 5 and over 5 days.

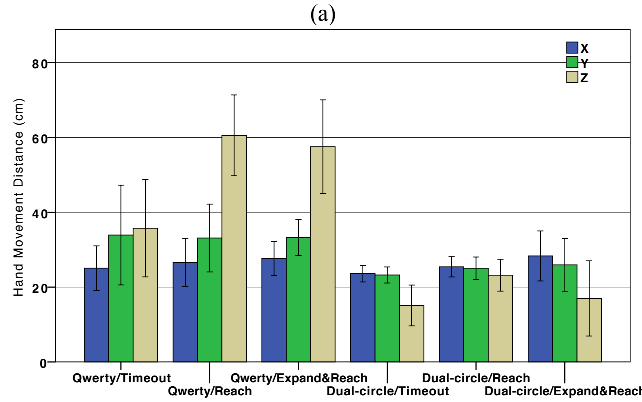
	Qwerty			Dual-circle		
	Time-out	Reach	Expand &Reach	Time-out	Reach	Expand &Reach
Day 1	3.00	10.44	6.11	1.63	6.42	2.27
Day 5	1.61	6.46	1.60	0.00	5.81	1.38
5 days overall	1.60	6.24	4.14	0.64	5.45	2.58

5.3 Hand Movement in 3D Space

We also recorded the hand movement distance per character in day 5. A repeated-measures ANOVA for Layout x Selection Technique was used to analyze hand movement distance per character. Main effects were found for Layout ($F_{1,5}$=39.42, p<.01) and Selection Technique ($F_{2,10}$=10.21, p<.01). No interaction effect was found. Post hoc Bonferroni pairwise comparisons showed that Qwerty layout required significantly more hand movement distance than Dual-Circle (p<.01). The Timeout selection technique had significantly less movement distance than Reach (p<.05). There were no other significant differences between selection techniques. Mean hand movement distance per character is shown in Figure 6(a).



(a)



(b)

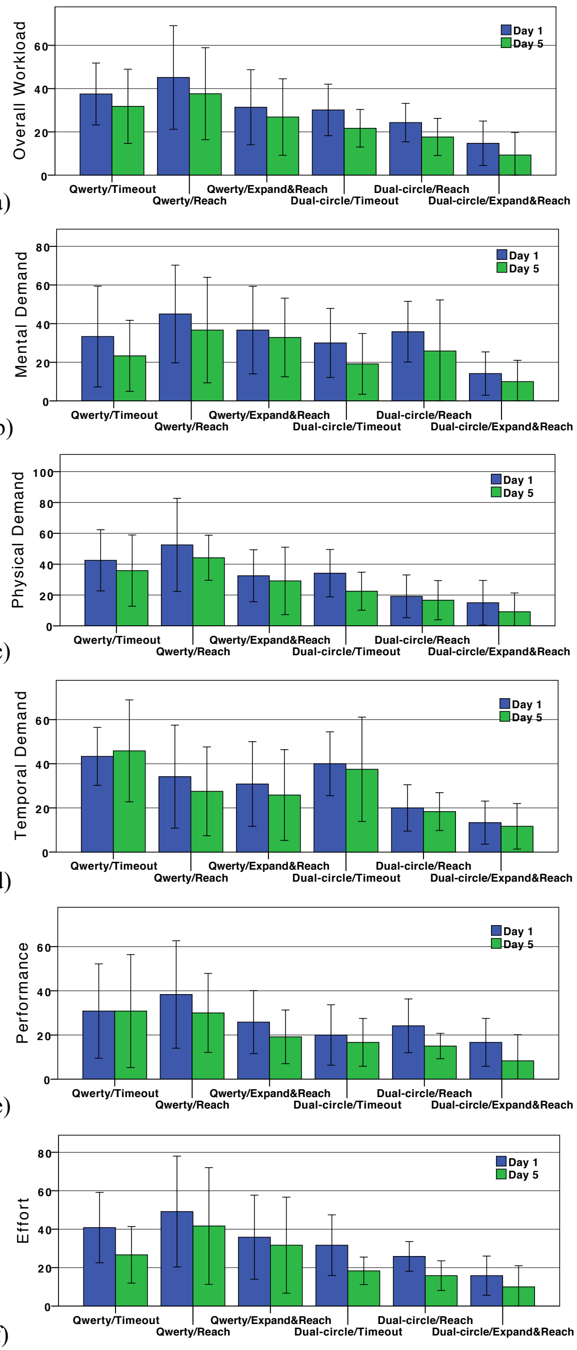
Figure 6. Mean hands movement distance per character. (a) Hand movement distance in 3D space (b) Hand movement distance in X, Y and Z axis.

We analyzed the hand movement distance in different axes (i.e. X, Y, Z) using one-way ANOVA for six text entry methods. We found that with Qwerty/Timeout and Dual-circle/Reach there was no significant effect of axis (p<.05). With Qwerty/Reach, Qwerty/Expand&Reach and Dual-circle/Timeout, main effects were found for Axis (p<.001). Post hoc Bonferroni pairwise comparisons showed that with Qwerty/Reach and Qwerty/Expand&Reach methods, hand movement in the Z axis was significantly more than in the X and Y axes (p<.05), while with Dual-circle/Timeout and Dual-circle/Expand&Reach methods, hand movement in the Z axis was significantly less than in the X and Y axes (p<.01). Mean hand movement distance per character in the X, Y and Z axes is shown in Figure 6(b).

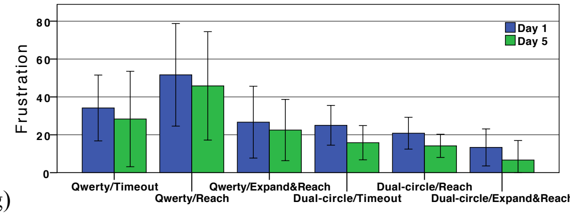
5.4 Task Load

A repeated-measures ANOVA for Layout x Selection Technique x Day was used to analyze the NASA task load index (TLX).

Main effects were found for Layout ($F_{1,5}$=9.21, p<.05), Selection Technique ($F_{2,10}$=5.75, p<.05) and Day ($F_{1,5}$=7.61, p<.05). No interaction effects were found. Post hoc Bonferroni pairwise comparisons showed that the Qwerty layout had significantly higher task load than the Dual-circle layout (p<.05), the Reach selection technique had significantly higher task load than Expand&Reach (p<.05), and the task load on day 1 was significantly higher than on day 5 (p<.05). Figure 7(a) shows the overall workload. Figures 7(b) to (g) show users' mental demand, physical demand, temporal demand, performance, effort and frustration.



(a) (b) (c) (d) (e) (f)



(g)

Figure 7. NASA task load index (TLX) results. (a) Overall workload (b) Mental demand (c) Physical demand (d) Temporal Demand (e) Performance (f) Effort (g) Frustration

5.5 User Preference

User preference data for each input method was collected on the first day and the last days. Users gave their preference (from 1 for strongly dislike to 10 for strongly like) after each text entry method, as shown in Figure 8. Overall, all participants preferred Dual-circle/Expand&Reach on both the first and last days of the study. The Dual-circle/Expand&Reach technique could enable people to input text comfortably in both "walk up and use" and the slightly longer term use (5 days) of our study. Its error tolerance also allows more casual hand movements without high concentration and physical effort.

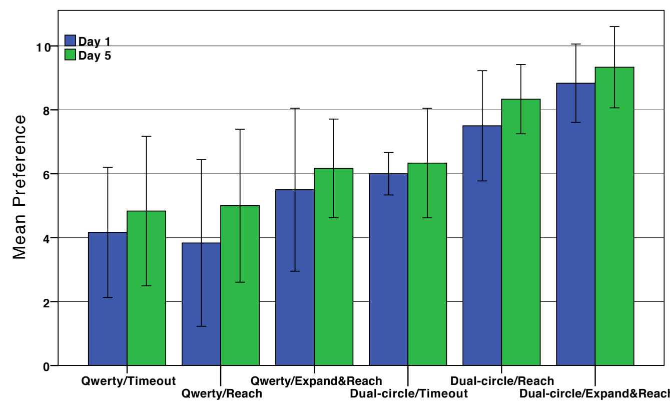


Figure 8. User preference

6. DISCUSSION

6.1 Text Entry Method

The dual-circle layout had better performance and lower task load than the Qwerty layout. Although users are more familiar with the Qwerty keyboard, its characters are packed in 2D, so character selection is more difficult than with the dual-circle layout. We noticed that when users select a character with one hand, they normally put down the other hand to avoid careless error selection. With the dual-circle layout, they can just relax the hand because there is enough blank space in the center to prevent careless error selection.

When using the Qwerty/Reach text entry method, users felt it was difficult to find the reach point in the beginning due to the absolute character position. High physical demand was also reported for Qwerty/Reach (Figure 7c). With Qwerty/Expand&Reach, on the other hand, as relative location was used, it was easier to select. However, both Qwerty/Reach and Qwerty/Expand&Reach required long movement distance when inputting text (Figure 6a), which was largely due to the hand movement for character selection along the Z dimension (Figure 6b). In contrast, as no hand movement along the Z dimension was required with the dual-circle layout, and users felt more easy and relaxed in their text entry tasks.

Participants also noticed their improvement for the Reach and Expand&Reach techniques with both keyboard layouts. Especially with the Dual-circle layout, typing speed increased continuously over 5 days (Figure 4). Timeout selection had few errors with both layouts, but the dwell time slowed text entry speed so it was not liked by users. Typing speed with the timeout method, with both layouts, did not improve in the course of the study as other text entry methods did.

Overall, Dual-circle/Expand&Reach was the best text input method. It is error tolerant thanks to the Expand&Reach selection technique so users make fewer careless wrong selections. The dual-circle layout also leaves enough blank space in the center to help people to relax their hands when not inputting text without worrying about triggering a selection when relaxing the hand. Thus Dual-circle/Expand&Reach is a practical text entry method for interactive TV. The Dual-circle/Reach method is more straightforward for selection but was found to be error-prone in this study. It could be a fast entry method given future improvements in tracking resolution. Qwerty/Expand&Reach could be used in scenarios in which familiarity with the standard Qwerty keyboard is desirable. The Qwerty/Reach method, however, may be a less promising method, given its high error rate and the high mental and physical effort noted by the participants.

6.2 Timeout Dwell Time

Typing speed with the timeout method, with both layouts, did not improve with practice. As noted above, the 1.2 s dwell time was based on a pilot study that showed a dwell time of less than 1.2 s produced more errors for beginners, and the observation that most commercial Kinect timeout based interaction uses more than 1.5 s. Our results suggest that the 1.2 s dwell time was suitable for first time use, however, it could be unnecessarily long for more experienced users, thus limiting performance. We also found that error rates using the timeout selection technique were low. With Dual-circle/Timeout, there were no errors at all in the last two days (Figure 5). Users also commented that the dwell time for the Dual-circle/Timeout method could be shorter after some practice.

For the dual-circle layout, it is possible to reduce the dwell time to any value. For example, the dwell time could be reduced even to zero, and in this extreme case, the Dual-circle/Timeout method will be equivalent to the Dual-circle/Reach method, leading to faster typing speed but more errors. With the Qwerty layout, however, the dwell time cannot be reduced so much. This is because users must move their hands over other characters to select the desired character in the Qwerty layout, thus a very small dwell time would trigger the undesired selection easily and lead to a huge increase in errors.

In a previous study using eye gaze for text entry, Majaranta et al. [23] showed that adjustable dwell time could improve text entry performance with a Qwerty layout keyboard, and this could also be the case for freehand text entry. However, in their experiment, participants used a 282 ms dwell time in the last session [23], which may be too short for a Qwerty layout with freehand gestural input. Further investigations of optimal dwell times for freehand interaction could form part of future work.

6.3 Tracking Sensors and Freehand Gestural Text Input

Our work aims to facilitate freehand gestural interaction tracked by a low-cost remote single camera with no requirement for a user to carry, wear or pick up an input device or fiducial markers. We used Microsoft Kinect in this study as it is a typical currently available low-cost remote gesture tracking sensor. Some more accurate tracking systems, such as Vicon cameras[4], can offer high accuracy but they require markers on the body and calibration before use, which we explicitly want to avoid. Other sensors, such as the soon to be available LeapMotion sensor[5], may provide accurate hand tracking without markers, but the tracking range is relatively short and thus may not suitable for interactive TV.

Even with remote single tracking sensors having higher resolution and become more affordable, there are still some common limitations of freehand interaction that cannot be solved by increasing tracking resolution: e.g. (i) having only one tracking angle brings occlusion of finger/wrist motion, (ii) previous work has shown that when the hands are moving freely in 3D, the user cannot point as accurately as with a 2D surface [36], and (iii) motor control with large muscle groups, e.g. the shoulder, is less accurate than with small muscle groups, e.g. fingers [10, 32]. Even with technical advances, remote camera tracking will share many of the current benefits and limitations and our main findings are broadly applicable across this class of devices. It is highly unlikely that improvements in tracking resolution would have a detrimental effect on Expand&Reach. Dual-circle/Reach, as suggested above, although error-prone in this study, could be a fast entry method given future improvements in tracking accuracy.

The performance of Dual-circle/Expand&Reach (5.56 wpm on the first day, 8.46 wpm on the last day and a mean of 7.01 wpm over 5 days) also compares well to similar gestural text input without word prediction, such as a mean of 5.4 wpm after 4 days' practice using accelerometer sensors in hand held devices [17], a mean of 6.5 wpm over 2 weeks with a data glove and fiducial trackers [25], and a mean of 5.18 wpm with a text entry method combing speech and gesture [13]. The performance of Dual-circle/Expand&Reach is also much better than the 1.83 wpm text entry speed reported with the default Xbox gestural text input [13]. On the other hand, the performance of Dual-circle/Expand&Reach achieved in our study still cannot compete with some text entry methods enabled by more accurate tracking devices or input models, such as a Qwerty virtual keyboard and handheld devices (18.9 wpm) [34], or text entry using gaze (nearly 20 wpm) [23].

6.4 Design Suggestions

Interestingly, previous work using hand held devices for text input showed that the Qwerty layout was faster and had fewer errors compared to a circle layout when the users were about 2.5 m away from the display [34]. In our study, on the other hand, the results suggested exactly the opposite. Such differences are largely due to differences in user behavior with different input methods. Moving the bare hand and arm freely in the air is different from holding and moving a mouse on the desktop, tapping the thumb on a mobile phone, or multi-touch movements with the fingers on a touch sensitive surface. For example, in [34], the desired character is selected by pointing and pressing a button using a Wii

[4] http://www.vicon.com/products/cameras.html

[5] https://www.leapmotion.com/

remote, however, in freehand gestural interaction button pressing is not available and so must be replaced by other techniques, such as Timeout, Reach, and Expand&Reach.

From the findings of our study, some design guidelines emerge for freehand interaction.

(1) Although the hand can move freely in 3D space, and hand motion in the Z dimension could be used as a gesture delimiter or trigger [32], frequent movements in the Z dimension are not recommended for freehand gestural interaction. This is not only because the hand moves more slowly forward and backward [11], and actions in the Z dimension can be error-prone [32], but also because frequent movements in the Z dimension can increase the hand movement distance and corresponding physical demands, as shown in this study.

(2) The circle layout is useful for interface design with freehand gesture. In the circle layout, all items have one side facing to the center and, therefore, can be reached directly without moving the hands over other objects. And the large blank area in the center can allow users to relax their hands without worrying about triggering undesired actions, thus potentially reducing arm fatigue.

(3) Combing the expanding target and reach selection techniques is useful for freehand interaction, especially for hand motion tracking by inexpensive sensors with low resolution. When used with the reach technique, the target can expand not only in virtual space but also in motor space. Thus, using the expanding target and reach techniques together addresses the limitation that targets can expand only in virtual space [24], and it is also well suited to freehand interaction techniques due to its error tolerance.

7. CONCLUSIONS

Freehand gesture is a promising interaction technique for interactive TV. The work presented here reports an initial investigation of the design and evaluation of potential text entry methods. The designs proposed in this paper, such as the circle layout, expanding target and the reach selection technique, are not entirely new individually. But their combination and their use with freehand gesture have never previously been explored. This novel combination of layouts familiar to Qwerty users and selection techniques designed specifically for freehand gestural text input are well suited to the intended interactive TV context, and we have demonstrated the effectiveness of this combination. Future work includes integration of word-level functions such as word prediction and completion, as well as evaluation of emerging more advanced tracking sensors.

8. ACKNOWLEDGMENTS

This research was supported in part by a Google Research Award.

9. REFERENCES

[1] J. Accot and S. Zhai. More than dotting the i's --- foundations for crossing-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, 73--80.

[2] R. Aoki, B. Chan, M. Ihara, T. Kobayashi, M. Kobayashi, and S. Kagami. A gesture recognition algorithm for vision-based unicursal gesture interfaces. In *Proceedings of the 10th European Conference on Interactive TV and Video*, EuroiTV '12, 53--56, 2012. ACM.

[3] H. Benko. Beyond flat surface computing: challenges of depth-aware and curved interfaces. In *Proceedings of Multimedia 2012*, MM '12, 935--944.

[4] X. Bi, B. A. Smith, and S. Zhai. Quasi-qwerty soft keyboard optimization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, 283--286, 2010. ACM.

[5] J. Bobeth, S. Schmehl, E. Kruijff, S. Deutsch, and M. Tscheligi. Evaluating performance and acceptance of older adults using freehand gestures for tv menu control. In *Proceedings of the 10th European Conference on Interactive TV and Video*, EuroiTV '12, 35--44, 2012. ACM.

[6] S. J. Castellucci and I. S. MacKenzie. Gathering text entry metrics on android devices. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, 1507--1512, 2011. ACM.

[7] D. Chatterjee, A. Sinha, A. Pal, and A. Basu. An iterative methodolgy to improve TV onscreen keyboard layout design through evaluation of user studies. *Advances in Computing*, 2(5):81-91, 2012.

[8] M. D. Dunlop, N. Durga, S. Motaparti, P. Dona, and V. Medapuram. Qwerth: an optimized semi-ambiguous keyboard design. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile devices and Services*, MobileHCI '12, 23--28, 2012. ACM.

[9] G. Geleijnse, D. Aliakseyeu, and E. Sarroukh. Comparing text entry methods for interactive television applications. In *Proceedings of the 7th European Conference on Interactive TV and Video*, EuroiTV '09, 145--148, 2009. ACM.

[11] T. Grossman and R. Balakrishnan. Pointing at trivariate targets in 3d environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, 447--454, 2004. ACM.

[12] F. Guimbretiére and T. Winograd. Flowmenu: combining command, text, and data entry. In *Proceedings of the 13rd Annual ACM Symposium on User Interface Software and Technology*, UIST '00, 213--216.

[13] L. Hoste, B. Dumas, and B. Signer. Speeg: a multimodal speech- and gesture-based text input solution. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, 156--163, 2012. ACM.

[14] A. Iatrino and S. Modeo. Text editing in digital terrestrial television: a comparison of three interfaces. In *Proceedings of the 4th European Conference on Interactive TV and Video*, EuroiTV '06.

[15] P. Isokoski and R. Raisamo. Quikwriting as a multi-device text entry method. In *Proceedings of the Third Nordic Conference on Human-Computer Interaction*, NordiCHI '04, 105--108, 2004. ACM.

[16] H.-C. Jetter, J. Gerken, and H. Reiterer. Why we need better modelworlds, not better gestures. In *Natural User Interfaces (Workshop CHI 2010)*, 1-4.

[17] E. Jones, J. Alexander, A. Andreou, P. Irani, and S. Subramanian. Gestext: accelerometer-based gestural text-entry systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, 2173--2182.

[18] M. Karam and m. c. schraefel. A taxonomy of gestures in human computer interactions. Technical report, University of Southampton, 2005.

[19] P. O. Kristensson, T. Nicholson, and A. Quigley. Continuous recognition of one-handed and two-handed gestures using 3d

full-body motion tracking sensors. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, IUI '12, 89--92, 2012. ACM.

[20] P. O. Kristensson and S. Zhai. Shark2: a large vocabulary shorthand writing system for pen-based computers. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, 43--52.

[21] F. C. Y. Li, R. T. Guy, K. Yatani, and K. N. Truong. The 1line keyboard: a qwerty layout in a single line. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, 461--470, 2011. ACM.

[22] I. S. MacKenzie and R. W. Soukoreff. Phrase sets for evaluating text entry techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, 754--755, 2003. ACM.

[23] P. Majaranta, U.-K. Ahola, and O. \v{S}pakov. Fast gaze typing with an adjustable dwell time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, 357--360, 2009. ACM.

[24] M. J. McGuffin and R. Balakrishnan. Fitts' law and expanding targets: experimental studies and designs for user interfaces. *ACM TOCHI*, 12(4):388--422, 2005.

[25] T. Ni, D. Bowman, and C. North. Airstroke: bringing unistroke text entry to freehand gesture interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, 2473--2476.

[26] D. A. Norman. Natural user interfaces are not natural. *Interactions*, 17(3):6--10, May 2010.

[27] D. A. Norman and J. Nielsen. Gestural interfaces: a step backward in usability. *Interactions*, 17(5):46--49, September 2010.

[28] K. Perlin. Quikwriting: continuous stylus-based text entry. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology*, UIST '98, 215--216, 1998. ACM.

[29] O. Polacek, M. Klima, A. J. Sporka, P. Zak, M. Hradis, P. Zemcik, and V. Prochazka. A comparative study on distant free-hand pointing. In *Proceedings of the 10th European Conference on Interactive TV and Video*, EuroiTV '12, 139--142, 2012. ACM.

[30] Primesense. The primesense 3D awareness sensor. .

[31] G. Ren and E. O'Neill. 3D marking menu selection with freehand gestures. In *IEEE Symposium on 3D User Interfaces*, 3DUI '12, 61-68.

[32] G. Ren and E. O'Neill. 3D selection with freehand gesture. *Computers & Graphics*, 2013.

[33] J. Rick. Performance optimizations of virtual keyboards for stroke-based text entry on a touch-based tabletop. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, 77--86, 2010. ACM.

[34] G. Shoemaker, L. Findlater, J. Q. Dawson, and K. S. Booth. Mid-air text input techniques for very large wall displays. In *Proceedings of Graphics Interface 2009*, GI '09, 231--238.

[35] A. J. Sporka, O. Polacek, and P. Slavik. Comparison of two text entry methods on interactive tv. In *Proceedings of the 10th European Conference on Interactive TV and Video*, EuroiTV '12, 49--52, 2012. ACM.

[36] R. J. Teather and W. Stuerzlinger. Assessing the effects of orientation and device on (constrained) 3d movement techniques. In *IEEE Symposium on 3D User Interfaces*, 3DUI '08, 43--50.

[37] R.-D. Vatavu. Interfaces that should feel right: natural interaction with multimedia information. In M. Grgic, K. Delac, and M. Ghanbari, editors, *Recent advances in multimedia signal processing and communications*, volume 231, page 145-170. Springer Berlin / Heidelberg, 2009.

[38] R.-D. Vatavu. User-defined gestures for free-hand tv control. In *Proceedings of the 10th European Conference on Interactive TV and Video*, EuroiTV '12, 45--48, 2012. ACM.

[39] A. Wexelblat. Research challenges in gesture: open issues and unsolved problems. In I. Wachsmuth and M. Fröhlich, editors, *Gesture and sign language in human-computer interaction*, volume 1371, page 1-11. Springer Berlin / Heidelberg, 1998.

[40] A. D. Wilson. Using a depth camera as a touch sensor. In *Proceedings of ACM International Conference on Interactive Tabletops and Surfaces*, ITS 2010, 69--72.

[41] J. O. Wobbrock, B. A. Myers, and J. A. Kembel. Edgewrite: a stylus-based text entry method designed for high accuracy and stability of motion. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, UIST '03, 61--70, 2003. ACM.

[42] S. Zhai and P. O. Kristensson. Interlaced qwerty: accommodating ease of visual search and input flexibility in shape writing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, 593--596, 2008. ACM.

[43] S. Zhai and P.-O. Kristensson. Shorthand writing on stylus keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, 97--104.